Operations Research and Fuzziology 运筹与模糊学, 2024, 14(3), 1221-1234
Published Online June 2024 in Hans. https://www.hanspub.org/journal/orf
https://doi.org/10.12677/orf.2024.143352

Hans汉斯

基于矩阵费雪分布的三维人脸变形模型

房　蔚

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2024年4月1日；录用日期：2024年6月23日；发布日期：2024年6月30日

摘　要

三维人脸的精确表示有利于各种计算机视觉和图形应用。然而，由于数据离散化和模型线性化，在目前的研究中获取准确的身份和表情线索仍然具有挑战性。本文提出了一种新的三维可变形人脸模型，来学习具有隐式神经表示的非线性连续空间。它构建了两个明确的解纠缠变形场来分别建模与身份和表情相关联的复杂形状，并且引入了一个神经混合场，自适应地混合一系列局部场来学习复杂的细节。其次，我们发现姿态参数在网络中可以更好地被解纠缠，对于人脸变形过程中发生的姿态变换，我们利用基于旋转矩阵的费雪分布矩阵来表示人脸姿态的角度，并模拟头部旋转的不确定性。实验表明我们的方法在人脸细节建模和姿态估计方面具有优越性。

关键词

三维可变形人脸模型，隐式神经表示，姿态估计，矩阵的费雪分布

3D Face Morphable Model Based on Matrix Fisher Distribution

Wei Fang

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 1st, 2024; accepted: Jun. 23rd, 2024; published: Jun. 30th, 2024

Abstract

The accurate representation of 3D faces is beneficial to various computer vision and graphics applications. However, due to data discretization and model linearity, it is still challenging to obtain accurate identity and expression cues in current research. In this paper, we propose a new 3D deformable face model to learn a nonlinear continuous space with implicit neural representations. It constructs two explicit disentanglement deformation fields to model the complex shapes asso-

文章引用: 房蔚. 基于矩阵费雪分布的三维人脸变形模型[J]. 运筹与模糊学, 2024, 14(3): 1221-1234.
DOI: 10.12677/orf.2024.143352

ciated with identity and expression respectively, and introduces a neural hybrid field to learn complex details by adaptively mixing a series of local fields. Secondly, we find that the pose parameters can be better disentangled in the network. For the pose transformation during face deformation, we use the Fisher distribution matrix based on the rotation matrix to represent the angle of the face pose and simulate the uncertainty of the head rotation. Experiments show that our method has advantages in face detail modeling and pose estimation.

Keywords

3D Face Morphable Model, Implicit Neural Representations, Pose Estimation, Fisher Distribution of Matrix

1. 引言

三维形变人脸模型(3D Face Morphable Model, 3DMM) [1]是一个著名的统计模型，通过学习一组具有密集对应关系的面部形状和纹理的先验分布而建立的，旨在渲染具有高度多样性的真实人脸。3DMM 在计算机视觉、计算机图形学、生物识别学和医学成像领域的许多人脸分析应用中被广泛利用。

在 3DMM 中，最基本的问题在于如何生成潜在的变形表示，在过去的二十年里，随着数据在规模性、多样性和质量方面的改进，重建方法取得了显著的进展。这些方法最初是基于线性模型[2]的，并进一步扩展到多线性模型[3]，其中不同的模式单独编码。然而，由于线性模型的表示能力相对有限，这些方法并不能很好地处理复杂变化的情况，如夸张的表情变化。在深度学习的背景下，通过使用卷积神经网络(CNN)或图神经网络，以 2D 图像[4]或 3D [5]网格为输入，研究了许多非线性模型，他们确实带来了一定的提高。然而，受输入数据上离散表示方法的分辨率限制，人脸先验信息没有被充分捕获，导致形状细节会丢失。此外，目前所有的方法都依赖于点对点对应[6]的预处理过程，但是人脸配准问题本身仍然具有挑战性。

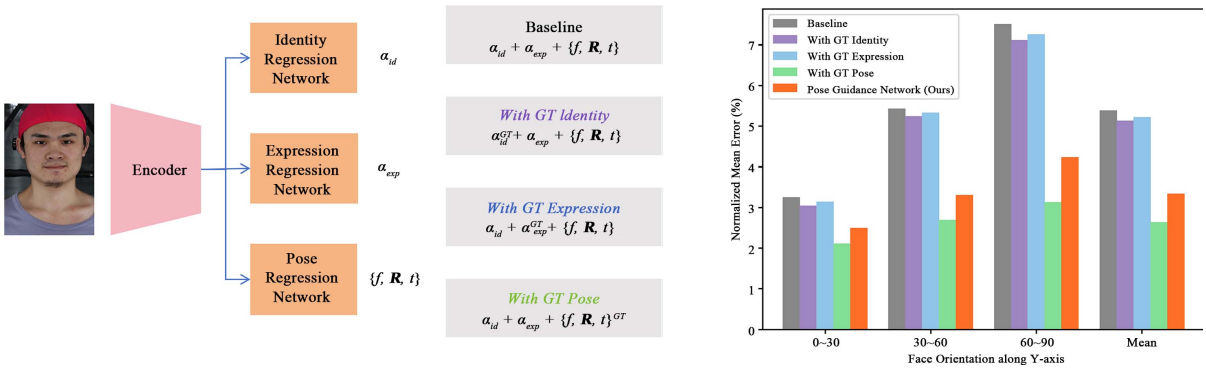

Figure 1. Parameter regression results
图 1. 参数回归结果

最近，关于隐神经表示(Implicit Neural Representations, INRs) [7]的一些研究表明，通过学习连续的深度隐式函数可以精确地建模三维几何。它们将一个输入观测描述为一个低维的形状嵌入，并估计一个查

询点的有符号距离函数(Signed Distance Function, SDF)或占有率值，从而可以通过等照度定义任意分辨率和拓扑结构的曲面。由于参数的连续化和一致的表示，INRs 优于离散的体素、点云和网格，并在形状重建[8]和表面配准[9]中得到了良好的实验结果。这种优势表明它已经渐渐成为一种替代传统 3DMM 的重建方法。然而，与室内场景或人体等具有明显形状差异和有限非刚性变化的物体不同，所有面部表面看起来非常相似，包含更复杂的变形，并且多种身份和丰富的表情会交织在一起，这使得目前的 INRs 方法在人脸建模方面存在问题。

另一方面，传统 3DMM 即参数面模型通常包含三组参数：身份、表情和姿态。基于 CNN 的方法直接学习了三维人脸模型的参数的回归，达到了最先进的性能。本文对 FaceScape [10]数据集进行了详细的调查，以评估这些参数在 CNN 中能否被很好地解纠缠以及准确地回归。结果见图 1。首先训练一个神经网络，它以一个 RGB 图像作为输入，同时回归身份、表情和姿态参数。基线 3DMM 模型是通过最小化三维顶点误差，即归一化平均误差(NME)得到的，然后，分别将预测的身份、表情和位姿参数替换为相应的地面真实参数(记为 GT 身份、GT 表情和 GT 姿态)，并重新计算三维人脸重建误差。令人惊讶的是，本文发现 GT 姿态的性能增益几乎是其两个同类产品的 2 倍。当面部取向程度的增加时，改善更加显著。假设导致这一结果的原因有两个：1) 这三组参数高度相关，预测一个不良姿态将很大程度影响三维人脸模型的身份和表情估计；2) 三维面部标注很少，特别是对于那些具有不寻常的姿态。

因此，一个需要解决的问题是如何有效地表示头部姿势的角度，在以往的研究中[11]，经常采用欧拉角和单位四元数这两种表示形式，使用欧拉角会导致被称为框架锁[12]的模糊性问题。另一种表示是单位四元数，由于它的双重嵌入，可能导致两个不连通局部最小值存在，因此也存在模糊问题。此外，Zhou 等人[13]证明了任何四维或更少维度的旋转表示都是不连续的，这对训练网络的优化是一个麻烦。为了克服模糊性问题，本研究采用了旋转矩阵表示法。基于图像的旋转矩阵表示的姿态估计任务等价于 $SO(3)$ 估计旋转矩阵。然而，由于头部姿态的不确定性，旋转矩阵很难直接预测。为了处理姿态的不确定性，在方向统计量[14]中定义了旋转的各种概率分布。矩阵费雪分布是在[15]中引入的旋转矩阵的指数概率密度。$SO(3)$上的矩阵费雪分布可以由 9 个参数来定义，它对应于 $\mathbb{R}^3$ 中由三维均值和六维协方差来定义的高斯分布。$SO(3)$上的矩阵费雪分布可以约束旋转矩阵，并且为进一步的分析和估计提供了不确定性。

为了解决上述问题，本文提出了一种新颖的 3D 人脸形变模型，它通过学习 INRs 对传统的 3DMMs 进行了实质性的升级。为了捕获非线性的面部几何变化，该方法构建了独立的隐式神经子网络，将形状变形显式地解耦为身份和表情的两个变形场。对于姿态的变化，本章利用基于旋转矩阵的费雪分布矩阵来表示人脸姿态的角度，并模拟头部旋转的不确定性。

2. 相关工作

2.1. 三维人脸变形模型

作为一种通用的人脸表示方法，通过使用非刚性迭代最近点算法[16]将已知模板网格注册到所有训练扫描中，并使用主成分分析方法跨越先验人脸分布。为了对身份依赖的表情进行建模，3DMM 被进一步扩展到多线性模型[17]。在此之后，数据改进[18]取得了巨大的进步，FLAME [19]是一种结合颌骨关节和线性表情混合形状来控制面部表情的人脸模型，它从一个包括 D3DFACS [20]的大型 3D 人脸数据集中学习，并提供了比以往更令人印象深刻的结果，然而它的非线性面部变形不能被很好地捕获。

2.2. 隐式神经表示

近年来，隐式神经表示已经成为一种更有效、更合适的三维几何表示，因为它们不能连续建模离散的形状。为了保持细节，进一步利用了由形状元素[21]、网格[22]或八叉树[23]划分的结构化局部特征。

此外，为了很好地捕捉形状变化和对应关系，特别学习了一个额外的隐式变形潜在空间[24]。然而，现有的技术很难同时实现较高的视觉保真度和多样性，因此它们不适合进行可变形的人脸建模。此外，目前对 INRs 的研究主要集中在水密输入上，如 ShapeNet [25]中的输入，其中，H3D-Net [26]学习了一个隐式的头部形状空间用于 2D 重建；i3DMM [27]是第一个为人类头部设计的隐式 3D 可变形模型，然而，它在表示面部区域时严重受到低质量的影响。

2.3. 旋转矩阵表示

在姿态估计中，确定姿态的表示是非常重要的。一些使用深度学习模型的研究将欧拉角[28]或单位四元数[29]作为位姿估计的表示。然而，由于它们的性质，这些表示法存在模糊性问题。因此，旋转矩阵被用于姿态估计来解决这个问题。Prokudin 等人[30]探索了旋转矩阵结合奇异值分解进行姿态估计。Lee 等人[31]采用矩阵 Fisher 分布来表示姿态估计的不确定性，并根据贝叶斯框架构造了一个姿态估计器。Mohlin 等人[32]提出了一种近似的矩阵费雪分布方法的归一化常数来估计目标方向。Wang 等人[33]将矩阵费雪分布与高斯分布相结合，表示姿态存在较大的不确定性。

3. 方法

3.1. 人脸几何变形场

本文利用隐式神经表示来学习非线性三维变形人脸模型。所提出的方法明确地将面部形状变形分别分解为两个与身份和表情相关的独立变形场，并学习了一个深度 SDF 来表示模板形状。所有的变形场都与一系列的局部隐式函数混合，以进行更详细的表示。

它的基本思想是训练一个神经网络来拟合一个连续的函数 f，该函数通过水平集隐式地表示曲面，并且可以以各种格式进行去定义，例如占用率、SDF 或 UDF。该方法利用基于表情和身份的潜在嵌入的深度 SDF 来进行全面的人脸表示。它从一个查询点输出有符号的距离 s：

$$f:\left(\mathbf{p},\mathbf{z}_{\exp},\mathbf{z}_{id}\right)\in\mathbb{R}^3\times\mathbb{R}^{d_{\exp}}\times\mathbb{R}^{d_{id}}\to s\in\mathbb{R} \tag{1}$$

其中，$\mathbf{p}\in\mathbb{R}^3$ 为三维空间中查询点的坐标，$\mathbf{z}_{\exp}$ 和 $\mathbf{z}_{id}$ 分别表示表情和身份嵌入。

该方法的目标是学习一个神经网络来参数化，使它满足真正的面部形状先验。流程图见图 2，所提出的网络由三个形变网块组成，明确地解开了面部形状变形的学习过程，确保了个体间的差异和细粒度变形的准确建模。特别地，前两个形变块分别学习与表情和身份变化相关的单独变形场，而模板形变网块学习模板面形状的有符号距离场。

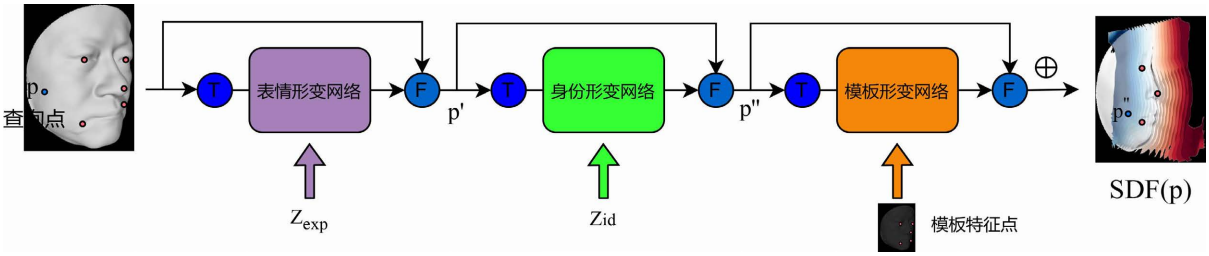

Figure2. Algorithm flow chart
图 2. 算法流程图

上面所有变形场都是由共享的形变网块架构实现的，其中整个面部变形或几何进一步分解为许多语义有意义的部分，并由一组局部函数编码，从而可以充分捕获丰富的细节。一个以查询点位置为条件的

轻量级模块，即融合网络，被堆叠在形变网块的末端，以自适应地混合局部场。因此，实现了一个复杂的神经混合场。本章工作的三个核心组件和它们的结构都有相应的轻微变化。对它们的简要描述如下：

1) 表情形变网络(ExpNet)

由表情引起的面部变形用 ExpNet $\mathbf{E}$ 表示，SoftMax 在每次面部扫描中学习一个表情变形：

$$\mathbf{E}:\left(\mathbf{p},\mathbf{z}_{\exp},l\right)\to\mathbf{p}'\in\mathbb{R}^3 \tag{2}$$

其中，$l\in\mathbb{R}^{k\times3}$ 表示由特征点网络 $\eta:\left(\mathbf{z}_{\exp},\mathbf{z}_{id}\right)\to l$ 生成的观察面上的 k 个三维特征点，引入来定位神经混合场中的查询点 $\mathbf{p}$。观测空间中的点 $\mathbf{p}$ 被 $\mathbf{E}$ 变形为人特定的规范空间中的一个新的点 $\mathbf{p}'$，它表示具有中性表情的面。

2) 身份形变网络(IDNet)

为了模拟个体之间的形状变形，IDNet $\mathbf{I}$ 进一步将规范空间变形为一个由所有面共享的模板形状空间：

$$\mathbf{I}:\left(\mathbf{p}',\mathbf{z}_{id},l'\right)\to\left(\mathbf{p}'',\delta\right)\in\mathbb{R}^3\times\mathbb{R} \tag{3}$$

其中，$l'\in\mathbb{R}^{k\times3}$ 表示由另一个仅基于身份嵌入的特征点网络 $\eta':\mathbf{z}_{id}\to l'$ 生成的标准面上的 k 个特征点，$\mathbf{p}''$ 为模板空间中的变形点。为了处理预处理过程中可能产生的不存在的对应关系，另外预测了一个残差项 $\delta\in\mathbb{R}$ 来校正预测的 SDF 值 s_0。

3) 模板形变网络(TempNet)

TempNet $\mathbf{T}$ 学习共享模板面的一个有符号的距离场：

$$\mathbf{T}:\left(\mathbf{p}'',l''\right)\to s_0\in\mathbb{R} \tag{4}$$

其中，$l''\in\mathbb{R}^{k\times3}$ 为模板面上的 k 个特征点，在整个训练集上取平均值，s_0 为未校正的 SDF 值。一个查询点的最终 SDF 值是通过 $s=s_0+\delta$ 计算出来的，模型最终可以表述为：

$$f\left(\mathbf{p}\right)=\mathbf{T}\left(\mathbf{I}_{p''}\left(\mathbf{E}\left(\mathbf{p},\mathbf{z}_{\exp}\right),\mathbf{z}_{id}\right)\right)+\mathbf{I}_{\delta}\left(\mathbf{E}\left(\mathbf{p},\mathbf{z}_{\exp}\right),\mathbf{z}_{id}\right) \tag{5}$$

所提出的方法通过以细粒度和有意义的方式解纠缠变形场来学习面部变形，以确保能够准确地学习更多样和复杂的面部变形。

4) 神经混合场

形变网块是由 $\mathbf{E}$、$\mathbf{I}$ 和 $\mathbf{T}$ 三个子网络共享的公共体系结构。神经混合场学习一个连续场函数 $\psi:\left(x\in\mathbb{R}^3,f\left(x\right)\in\mathbb{R}^{512}\right)\to v$，以产生一个用于综合面表示的神经混合场。其中 $f\left(x\right)$ 表示人脸图像提取得到的相关特征。特别地，为了克服单个网络的有限表达能力，将人脸空间分解为一组具有语义意义的局部区域，并在混合之前单独学习 v (例如变形或有符号距离值)。这种设计受到最近隐式神经表示对人体研究的启发，该研究引入了线性混合蒙皮算法[34]，使网络可以从身体某个部位的单独变换中学习。为了更好地表示详细人脸表面，将原始线性混合蒙皮算法中的常数变换项替换为 $\psi_n\left(x-l_n\right)$，并将神经混合场定义为：

$$v=\psi\left(x\right)=\sum_{n=1}^{k}\omega_n\left(x\right)\psi_n\left(x-l_n\right) \tag{6}$$

其中，l_n 是描述第 n 个局部区域的参数，$\omega_n\left(x\right)$ 是第 n 个混合权值，$\psi_n\left(x-l_n\right)$ 是相应的局部域。通过这种方式，在一系列局部域上执行混合，而不是计算一些固定位置的输出值 v 的加权平均值，从而在处理复杂的局部特征时具有更强的表示能力。

具体来说，利用位于眼角、嘴角和鼻尖的 5 个特征点来描述局部区域 $\left(l_n\in\mathbb{R}^3\right)_{n=1}^{5}$，每个区域被分配一个带有正弦激活的微型 MLP 来生成局部场，记为 ψ_n。为了捕获高频局部变化，在坐标 $x-l_n$ 上利用正

弦位置编码 γ。在形变网块的最后，配置了一个基于输入绝对坐标 x 的轻量级融合网络，该网络由一个 3 层的 MLP 结合 Softmax 实现，用于预测混合权重 $\left(\omega_n \in \mathbb{R}^+\right)_{n=1}^5$。

3.2. 矩阵费雪分布的概率姿态估计

本文用一个 SE(3)场 $(R,t)\in\mathbb{R}^6$ 来表示变形，其中 $R\in\mathbb{R}^3$ 是一个表示螺旋轴和旋转角度的旋转向量，变形坐标 x' 可以用 $Rx+t$ 来计算。该方法利用 SE(3)来描述面部形状变形，因为它在处理下颌旋转方面具有优越的能力，并且比普通平移变形 $x'=x+t$ 对姿态扰动具有更好的鲁棒性。在本文中，首先提出了旋转矩阵作为人脸姿态的表示。然后引入基于旋转矩阵的矩阵费雪分布对人脸姿态旋转不确定度进行建模，并讨论了其归一化常数的计算方法。

3.2.1. 人脸姿态表示

介绍了以旋转矩阵作为头人脸姿态角度表示的假设。在三维世界坐标中，位于 (u,v,w) 处的点首先绕 x 轴旋转角度 α，然后绕 y 轴旋转角度 β，最后绕 z 轴旋转角度 χ。这样的旋转序列可以表达为：

$$R_z(\chi)R_y(\beta)R_x(\alpha)\cdot(u,v,w)^{\mathrm{T}}=R\cdot(u,v,w)^{\mathrm{T}} \tag{7}$$

其中 R 为旋转矩阵，可以表示从固定系到惯性坐标系的线性变换。人脸姿态的角度可以用旋转矩阵 R 表示，如下面公式所示：

$$R=\begin{bmatrix}\cos\beta\cos\chi & \sin\alpha\sin\beta\sin\chi-\cos\alpha\sin\chi & \cos\alpha\sin\beta\cos\chi+\sin\alpha\sin\chi\\ \cos\beta\sin\chi & \sin\alpha\sin\beta\sin\chi+\cos\alpha\sin\chi & \cos\alpha\sin\beta\cos\chi-\sin\alpha\sin\chi\\ -\sin\beta & \sin\alpha\cos\beta & \cos\alpha\cos\beta\end{bmatrix} \tag{8}$$

其中，χ、β 和 α 分别表示偏航角、俯仰角和横滚角。此外，针对现有的人脸姿态估计数据集，标签采用欧拉角表示，所以这些标签应该转化为旋转矩阵。欧拉角可以通过旋转矩阵的元素来确定，这种转化可以用表达式来表示：

$$\begin{cases}\alpha=a\tan 2\left(R_{32},R_{33}\right)\\ \beta=a\tan 2\left(-R_{31},\sqrt{R_{32}^2+R_{33}^2}\right)\\ \chi=a\tan 2\left(R_{21},R_{11}\right)\end{cases} \tag{9}$$

三维旋转群，通常记为 $SO(3)$，是围绕三维欧几里得空间原点的所有旋转的群。旋转矩阵是李群 $SO(3)$ 上的三维特殊正交矩阵，满足以下性质：

$$SO(3)=\left\{R\in\mathbb{R}^{3\times3}\middle|R^{\mathrm{T}}R=I_{3\times3},\det[R]=+1\right\} \tag{10}$$

其中，$\det[\cdot]$ 是一个方阵的行列式。人脸姿态估计可以设计在三维特殊正交组上，以避免奇异性和模糊性问题。旋转矩阵难以约束，容易失去其正交性，因此，为了在 $SO(3)$ 上构造确定性人脸姿态，本文在特殊正交群上构造了指数密度模型的一种紧凑形式，即矩阵费雪分布来约束旋转矩阵，并表示人脸姿态旋转不确定性。

3.2.2. 矩阵费雪分布

矩阵费雪分布可以用来处理旋转的统计量。当应用于三维特殊正交组时，可通过九个参数确定，并对头部姿态不确定性进行建模。对于旋转矩阵 R，矩阵费雪分布由下面的概率密度函数定义：

$$p(R|M)=\frac{1}{a(M)}\exp\left(tr\left[M^{\mathrm{T}}R\right]\right) \tag{11}$$

其中，$M\in\mathbb{R}^{3\times3}$ 是一个无约束矩阵，$a(M)\in\mathbb{R}$ 是分布的归一化常数，$tr[\cdot]$ 是该矩阵的迹。该分布可以用 $R\sim \mathrm{M}(M)$ 表示。

归一化常数 $a(M)$ 由下式给出：

$$a(M)=\int_{R\in SO(3)}\exp\left(tr\left[M^{\mathrm{T}}R\right]\right)\mathrm{d}R \tag{12}$$

这对于精确和有效的计算是相当重要的。为了求值 $a(M)$，本文考虑了 M 的奇异值。假设无约束矩阵 M 的奇异值分解由下式给出：

$$M=USV^{\mathrm{T}} \tag{13}$$

其中，$S=diag(s_1,s_2,s_3)$ 是 M 的奇异值的对角矩阵且 $s_1\geq s_2\geq s_3\geq 0$。$U$ 和 V 是正交矩阵，满足 $U^{\mathrm{T}}U=V^{\mathrm{T}}V=I_{3\times3}$。而在 $SO(3)$中它们不是旋转矩阵，因为 U 和 V 的行列式可以是−1。为了解决这个问题，采用如下变换：

$$\begin{cases}\hat{U}=Udiag\left(1,1,\det\left[U\right]\right)\\ \hat{S}=diag\left(s_1,s_2,s_3'\right)=diag\left(s_1,s_2,\det\left[UV\right]s_3\right)\\ \hat{V}=Vdiag\left(1,1,\det\left[V\right]\right)\end{cases} \tag{14}$$

$\hat{U}$ 和 $\hat{V}$ 的定义确保了 $\hat{U},\hat{V}\in SO(3)$。此外，$a(M)$ 的结果由 $\hat{S}$ 决定。接着，Λ 被定义为：

$$\Lambda=diag\left(s_1-s_2-s_3',s_2-s_1-s_3',s_3-s_1-s_2',\ s_1+s_2+s_3'\right) \tag{15}$$

归一化常数 $a(M)$ 可以表示为：

$$a(M)={}_1F_1(1/2,2,\Lambda) \tag{16}$$

其中，${}_1F_1(\cdot,\cdot,\cdot)$ 是广义超几何函数，要精确计算它是相当重要的。通过计算这个超几何函数，可以得到常数 $a(M)$ 及其梯度。

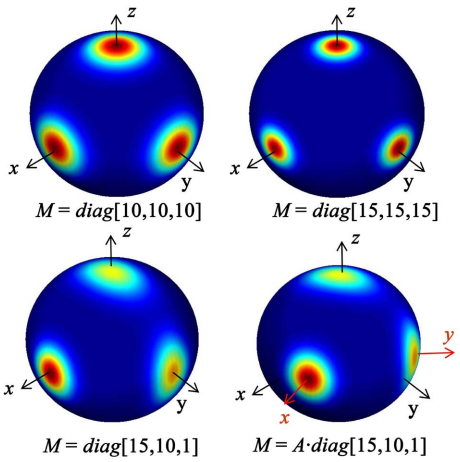

Figure 3. Visualization of matrix Fisher distribution with different M

图 3. 具有不同 M 的矩阵费雪分布的可视化图

当得到 $SO(3)$上的矩阵费雪分布时，在推断阶段，可以从该分布中提取出合适的旋转矩阵。假设 $R\in SO(3)$，且 $R\in \mathrm{M}(M)$，则最大均值可定义为：

$$R_{mean-\max}=\underset{R\in SO(3)}{\arg\max}\left\{p\left(R\middle|M\right)\right\} \tag{17}$$

计算之后，最优旋转 $\hat{R}\in SO(3)$ 表示为：

$$\hat{R}=R_{\text{mean-max}}=\hat{U}\hat{V}^{\mathrm{T}}=U\begin{bmatrix}1&0&0\\0&1&0\\0&0&\det[UV]\end{bmatrix}V^{\mathrm{T}} \tag{18}$$

为了可视化矩阵的费雪分布，提出了一种可视化 $SO(3)$上的概率密度函数的方法。首先计算旋转矩阵的第列向量的边际概率密度函数，并将其在单位球表面上可视化。矩阵费雪分布可视化见图 3。矩阵 M 决定了 $\mathrm{M}(M)$ 的形状。 M 的奇异值越大，边际概率密度的离散度越小。此外当 M 变化时，红色轴所示的主轴被旋转。

3.3. 损失函数

该模型使用多个损失函数进行训练，以学习合理的面部形状表示和稠密的对应关系。

3.3.1. 重建损失

基本的 SDF 结构损失被应用于学习隐式场：

$$\mathcal{L}_{sdf}^{i}=\lambda_1\sum_{\mathbf{p}\in\Omega_i}\left|f(\mathbf{p})-\overline{s}\right|+\lambda_2\sum_{\mathbf{p}\in\Omega_i}\left(1-\left\langle\nabla f(\mathbf{p}),\overline{n}\right\rangle\right) \tag{19}$$

其中，$\overline{s}$ 和 $\overline{n}$ 分别表示地面真实 SDF 值和场梯度。Ω_i 为人脸扫描第 i 个人的采样空间，λ 表示权衡参数。

3.3.2. Eikonal 损失

为了在整个网络中获得合理的隐式场，使用多个 Eikonal 损失来强制执行空间梯度的 $L-2$ 范数为单位 1：

$$\mathcal{L}_{eik}^{i}=\lambda_3\sum_{\mathbf{p}\in\Omega_i}\left(\left\|\nabla f(\mathbf{p})\right\|-1\right|+\left\|\nabla\mathbf{T}\left(\mathbf{I}(\mathrm{p}')\right)\right\|-1\right|\Big) \tag{20}$$

其中，$\mathcal{L}_{eik}^{i}$ 使网络在观测空间和正则空间中同时满足 Eikonal 损失。

3.3.3. 嵌入损失

它以零均值高斯先验对嵌入进行正则化：

$$\mathcal{L}_{emb}^{i}=\lambda_4\left(\left\|\mathbf{z}_{\exp}\right\|^2+\left\|\mathbf{z}_{id}\right\|^2\right) \tag{21}$$

3.3.4. 特征点损失

损失用于学习特征点网络：

$$\mathcal{L}_{lm}^{i}=\lambda_5\sum_{n=1}^{k}\left(\left|l_n-\overline{l_n^i}\right|+\left|l_n'-\overline{l_n'}\right|\right) \tag{22}$$

其中，$\overline{l^i}$ 表示样本 i 上的第 k 个标记的特征点，$\overline{l'}$ 表示对应的中性面上的特征点。

3.3.5. 特征点一致性损失

本文利用这一损失来引导变形的特征点位于地面真实中性面和模板面上的相应位置，以获得更好的对应性能：

$$\mathcal{L}_{lmc}^{i}=\lambda_6\sum_{n=1}^{k}\left(\left|\mathbf{E}(l_n)-\overline{l_n'}\right|+\left|\mathbf{I}\left(\mathbf{E}(l_n)-l_n''\right)\right|\right) \tag{23}$$

3.3.6. 负对数似然损失

给定一个样本(x, R_x)，x是输入图像，旋转矩阵$R_x \in SO(3)$是地面真实值。矩阵费雪分布模块的细节见图 4。通过将样本依次输入卷积神经网络和全连通层，得到M_x，然后生成矩阵费雪分布$\mathrm{M}(M_x)$。利用分布的负对数似然函数可以计算M_x和R_x之间的矩阵费雪分布损失(MFD 损失)。MFD 损失可以表示为：

$$\mathcal{L}_{MFD}^{i} = -\log\left(p\left(R_x \middle| M_x\right)\right) = \log\left(a\left(M_x\right)\right) - tr\left[M_x^{\mathrm{T}} R_X\right] \tag{24}$$

该损失是一个连续的凸函数，具有连续的梯度，适合于优化。

对所有以i为索引的人脸样本计算总的训练损失，最终表示为：

$$\mathcal{L} = \sum_i \left(\mathcal{L}_{sdf}^{i} + \mathcal{L}_{eik}^{i} + \mathcal{L}_{emb}^{i} + \mathcal{L}_{lm}^{i} + \mathcal{L}_{lmc}^{i} + \mathcal{L}_{MFD}^{i}\right) \tag{25}$$

4. 实验

4.1. 数据集

FaceScape [10]是一个大规模的高质量的 3D 人脸数据集，由 938 个个体组成，有 20 种表情类型。来自 365 个人的数据是公开的，本文主要用它们进行实验。具体来说，从 355 人的 15 个表情的 5323 次面部扫描作为训练集，从其余 10 人的 20 个表情的 200 次面部扫描作为测试集。

4.2. 网络结构

所有形变模块ψ_n被实现为具有 3 个隐藏层和 32 维正弦激活隐藏特征的 MLPs。超参网络ϕ_n是由 ReLU 激活的 3 层 MLPs，其中隐藏层维度为 64。特征点网络η和η'具有三个 128 维全连接层。

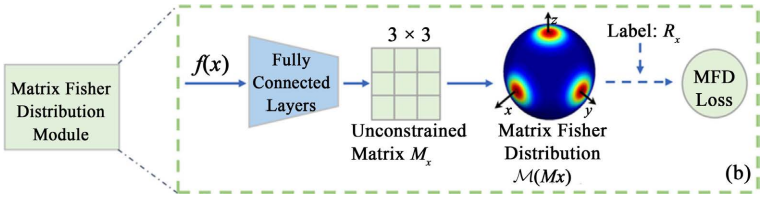

Figure 4. Matrix Fisher distribution module diagram
图 4. 矩阵费雪分布模块示意图

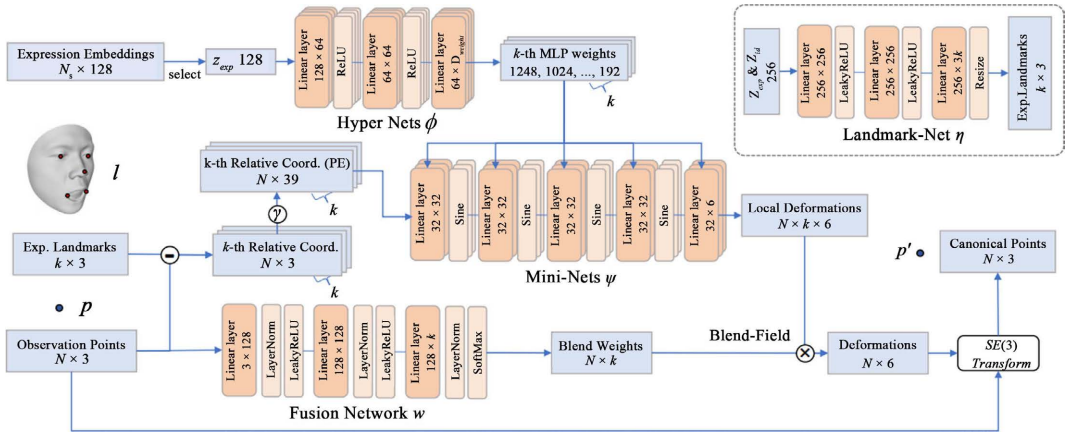

Figure 5. Expression deformation network structure
图 5. 表情形变网络框架结构

表情形变网络 **E** 的详细结构图见图 5，其中 N_s 表示人脸扫描总次数， N_{id} 表示对象身份数量，PE 表示位置编码。所有的网络都由 MLP 完全实现。为了获得更好的高频信息，本文通过正弦位置编码 γ 对 k 个特征点的相对坐标进行编码，写成：

$$\gamma(p)=\left(\sin\left(2^0\pi p\right),\cos\left(2^0\pi p\right),\cdots,\sin\left(2^{L-1}\pi p\right),\cos\left(2^{L-1}\pi p\right)\right) \tag{26}$$

在本文的实验中， $L=6$ ， $k=5$ 。此外，每个 ψ_n 都利用正弦激活，参数初始化。在训练过程中， k 个 ψ_n 的参数由 k 个对应的超参网络 ϕ_n 生成，以获得更具有表达性的潜在空间，这是最近隐式神经表示研究中常见的技术操作。训练网络的权衡参数 $\lambda_1,\lambda_2,\cdots,\lambda_7$ 分别设置为 3e3、1e2、5e1、1e6、1e2、1e2 和 1e2。

网络中的 SE(3)形变模块的输入 Deformations 即本文所描述的矩阵费雪分布模块所估计得到的人脸姿态，它的网络结构见图 4。

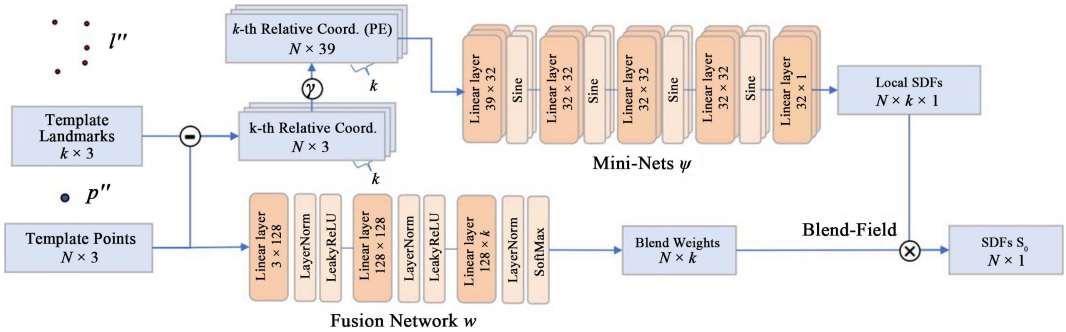

Figure 6. Templet deformation network structure
图 6. 模板形变网络结构

身份形变网络结构和表情形变网络结构相似，变形到模板空间中得到 $\mathbf{p}''$，模板形变网络结构见图 6，最后经过神经混合场得到 SDF s_0 。

4.3. 实验细节

该模型由 Adam 以端到端的方式进行训练。以初始学习率为 0.0001 训练模型 1500 个 epochs，在 200 个 epochs 之后，每 10 个 epochs 将其衰减 0.95 倍。在 1 块 NVIDIA RTX 3090 GPU 上进行训练，训练时间约为 2 天，微批大小为 72。在测试过程中，在单个 GPU 上优化 200 个样本大约需要 4 个小时。

4.4. 实验结果

本文使用提出的模型来拟合人脸扫描，并将重建结果与 FLAME [19]，i3DMM [27]和 ImFace [35]的几何模型进行了比较，表明了该方法的先进性。FLAME 代码用于拟合测试集中的全面扫描，有 300 个身份参数和 100 个表情参数。对于 ImFace，使用 FaceScape 发布的由 938 个个体建立的双线性模型进行测试，其中身份和表情参数分别为 300 和 52。测试扫描已经包含在 FaceScape 的训练集中。对于 i3DMM，由于原始模型只在 58 个个体上进行训练，因此在与本文相同的训练集上重新训练模型，以进行公平的比较。在 i3DMM 和本文方法中，身份和表情嵌入都是 128 维的。

4.4.1. 定性分析

可视化了不同模型所获得的重建结果见图 7，其中每一列对应一个具有非中性表情的测试人。研究结果还包括在学习过程中看不见的表情。i3DMM 是第一个针对人类头部的深度隐式模型，但在相对复杂

的情况下，它无法捕捉复杂的变形和细粒度的细节，从而导致重建的面孔上的伪影。FLAME 能够很好地呈现身份特征，但在处理非线性变形时，会产生僵硬的面部表情。ImFace 的表现更好，主要是由于高质量的训练扫描和测试面包含在训练集中，但它仍然不能精确地呈现表情形态。相比之下，本文方法重建了具有更准确的身份和表情特性的面孔，它能够通过更少的潜在参数来保持微妙和丰富的非线性面部肌肉变形，如皱眉和撅嘴。

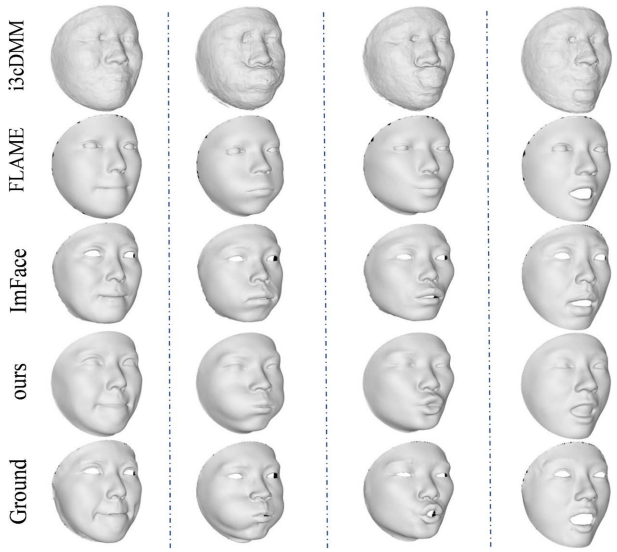

Figure 7. Reconstruction comparison results
图 7. 重建比较结果

4.4.2. 定量分析

具体地，采用对称倒角距离和 F-分数作为度量指标，并将 F 值的阈值设置为 0.001 作为严格标准，对称倒角距离越小效果越好，相反 F-分数越高效果越好。结果见表 1，可以看到，本文方法在两种度量指标下都超过了同类方法，这显然验证了它的有效性。

Table 1. Quantitative analysis comparison
表 1. 定量分析对比

Metrics	Dim.	Chamfer$(mm)^{\dagger}$	F-score@0.01$^{\diamond}$
i3DMM	256	1.635	42.26
FLAME	400	0.971	64.73
ImFace	352	0.625	91.11
our	256	**0.608**	**91.18**

4.4.3. 消融实验

本文方法建立在以下核心组件之上：人脸几何变形场，神经混合场。通过实验验证了这些设计的有效性。为了突出人脸几何变形场学习过程，建立了一个只包含其中一个变形场的基线网络，以普遍学习面部形状变形。因此，$\mathbf{z}_{\exp}$ 和 $\mathbf{z}_{id}$ 被连接为超参网络的输入。结果见表 2，定量结果表明了人脸几何变形场学习的意义。本文将 $\mathbf{E}$、$\mathbf{I}$、$\mathbf{T}$ 中的神经混合场替换为等量参数的 MLP，直接预测整个面部的全局变形或 SDF 值。表 2 中的定量结果证实了神经混合场在学习复杂表示方面的必要性。

Table 2. Ablation experimental results
表 2. 消融实验结果

Metrics	Chamfer$(mm)^{\dagger}$	F-score@0.01$^{\diamond}$
Ours w/o geo.	0.780	83.12
Ours w/o blend.	0.765	82.87
ours	**0.636**	90.95

5. 结论

本文提出了一种新颖的3D人脸形变模型，它通过学习INRs对传统的3DMMs进行了实质性的升级。为了捕获非线性的面部几何变化，该方法构建了独立的隐式神经子网络，将形状变形显式地解耦为身份和表情的两个变形场。对于姿态的变化，本文利用基于旋转矩阵的费雪分布矩阵来表示人脸姿态的角度，并模拟头部旋转的不确定性。

基金项目

国家自然科学基金资助项目(62273239)。

参考文献

[1] Blanz, V. and Vetter, T. (2023) A Morphable Model for the Synthesis of 3D Faces. In: Whitton, M.C., Ed., *Seminal Graphics Papers*: *Pushing the Boundaries*, *Volume* 2, Association for Computing Machinery, New York, 157-164. https://doi.org/10.1145/3596711.3596730

[2] Patel, A. and Smith, W.A.P. (2009) 3D Morphable Face Models Revisited. 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 20-25 June 2009, 1327-1334. https://doi.org/10.1109/cvprw.2009.5206522

[3] Bolkart, T. and Wuhrer, S. (2015) A Groupwise Multilinear Correspondence Optimization for 3D Faces. 2015 *IEEE International Conference on Computer Vision* (*ICCV*), Santiago, 7-13 December 2015, 3604-3612. https://doi.org/10.1109/iccv.2015.411

[4] Tran, L., Liu, F. and Liu, X. (2019) Towards High-Fidelity Nonlinear 3D Face Morphable Model. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 1126-1135. https://doi.org/10.1109/cvpr.2019.00122

[5] Ranjan, A., Bolkart, T., Sanyal, S., *et al*. (2018) Generating 3D Faces Using Convolutional Mesh Autoencoders. *Computer Vision-ECCV* 2018, Munich, 8-14 September 2018, 725-741. https://doi.org/10.1007/978-3-030-01219-9_43

[6] Liu, F., Tran, L. and Liu, X. (2019) 3D Face Modeling from Diverse Raw Scan Data. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 9407-9417. https://doi.org/10.1109/iccv.2019.00950

[7] Chen, Z. and Zhang, H. (2019) Learning Implicit Fields for Generative Shape Modeling. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 5932-5941. https://doi.org/10.1109/cvpr.2019.00609

[8] Genova, K., Cole, F., Vlasic, D., Sarna, A., Freeman, W. and Funkhouser, T. (2019) Learning Shape Templates with Structured Implicit Functions. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 7153-7163. https://doi.org/10.1109/iccv.2019.00725

[9] Liu, F. and Liu, X. (2020) Learning Implicit Functions for Topology-Varying Dense 3D Shape Correspondence. *Advances in Neural Information Processing Systems*, **33**, 4823-4834.

[10] Yang, H., Zhu, H., Wang, Y., Huang, M., Shen, Q., Yang, R., *et al*. (2020) FaceScape: A Large-Scale High Quality 3D Face Dataset and Detailed Riggable 3D Face Prediction. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 598-607. https://doi.org/10.1109/cvpr42600.2020.00068

[11] Ruiz, N., Chong, E. and Rehg, J.M. (2018) Fine-Grained Head Pose Estimation without Keypoints. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), Salt Lake City, 18-22 June 2018, 2155. https://doi.org/10.1109/cvprw.2018.00281

[12] Lepetit, V. and Fua, P. (2005) Monocular Model-Based 3D Tracking of Rigid Objects: A Survey. *Foundations and Trends® in Computer Graphics and Vision*, **1**, 1-89. https://doi.org/10.1561/0600000001

[13] Zhou, Y., Barnes, C., Lu, J., Yang, J. and Li, H. (2019) On the Continuity of Rotation Representations in Neural Networks. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 5738-5746. https://doi.org/10.1109/cvpr.2019.00589

[14] Mardia, K.V. and Jupp, P.E. (2009) Directional Statistics. John Wiley & Sons, Hoboken.

[15] Downs, T.D. (1972) Orientation Statistics. *Biometrika*, **59**, 665-676. https://doi.org/10.1093/biomet/59.3.665

[16] Amberg, B., Romdhani, S. and Vetter, T. (2007) Optimal Step Nonrigid ICP Algorithms for Surface Registration. 2007 *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, 17-22 June 2007, 1-8. https://doi.org/10.1109/cvpr.2007.383165

[17] Brunton, A., Bolkart, T. and Wuhrer, S. (2014) Multilinear Wavelets: A Statistical Shape Space for Human Faces. *Computer Vision-ECCV* 2014, Zurich, 6-12 September 2014, 297-312. https://doi.org/10.1007/978-3-319-10590-1_20

[18] Cao, C., Weng, Y., Zhou, S., Tong, Y. and Zhou, K. (2014) FaceWarehouse: A 3D Facial Expression Database for Visual Computing. *IEEE Transactions on Visualization and Computer Graphics*, **20**, 413-425. https://doi.org/10.1109/tvcg.2013.249

[19] Li, T., Bolkart, T., Black, M.J., Li, H. and Romero, J. (2017) Learning a Model of Facial Shape and Expression from 4D Scans. *ACM Transactions on Graphics*, **36**, Article No. 194. https://doi.org/10.1145/3130800.3130813

[20] Cosker, D., Krumhuber, E. and Hilton, A. (2011) A FACS Valid 3D Dynamic Action Unit Database with Applications to 3D Dynamic Morphable Facial Modeling. 2011 *International Conference on Computer Vision*, Barcelona, 6-13 November 2011, 2296-2303. https://doi.org/10.1109/iccv.2011.6126510

[21] Genova, K., Cole, F., Sud, A., Sarna, A. and Funkhouser, T. (2020) Local Deep Implicit Functions for 3D Shape. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 4856-4865. https://doi.org/10.1109/cvpr42600.2020.00491

[22] Ibing, M., Lim, I. and Kobbelt, L. (2021) 3D Shape Generation with Grid-Based Implicit Functions. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 13554-13563. https://doi.org/10.1109/cvpr46437.2021.01335

[23] Takikawa, T., Litalien, J., Yin, K., Kreis, K., Loop, C., Nowrouzezahrai, D., *et al*. (2021) Neural Geometric Level of Detail: Real-Time Rendering with Implicit 3D Shapes. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 11353-11362. https://doi.org/10.1109/cvpr46437.2021.01120

[24] Zheng, Z., Yu, T., Dai, Q. and Liu, Y. (2021) Deep Implicit Templates for 3D Shape Representation. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 1429-1439. https://doi.org/10.1109/cvpr46437.2021.00148

[25] Yi, L., Kim, V.G., Ceylan, D., Shen, I., Yan, M., Su, H., *et al*. (2016) A Scalable Active Framework for Region Annotation in 3D Shape Collections. *ACM Transactions on Graphics*, **35**, Article No. 210. https://doi.org/10.1145/2980179.2980238

[26] Ramon, E., Triginer, G., Escur, J., Pumarola, A., Garcia, J., Giro-i-Nieto, X., *et al*. (2021) H3D-Net: Few-Shot High-Fidelity 3D Head Reconstruction. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 5600-5609. https://doi.org/10.1109/iccv48922.2021.00557

[27] Yenamandra, T., Tewari, A., Bernard, F., Seidel, H., Elgharib, M., Cremers, D., *et al*. (2021) I3DMM: Deep Implicit 3D Morphable Model of Human Heads. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 12798-12808. https://doi.org/10.1109/cvpr46437.2021.01261

[28] Yuan, H., Li, M., Hou, J. and Xiao, J. (2020) Single Image-Based Head Pose Estimation with Spherical Parametrization and 3D Morphing. *Pattern Recognition*, **103**, Article 107316. https://doi.org/10.1016/j.patcog.2020.107316

[29] Hsu, H., Wu, T., Wan, S., Wong, W.H. and Lee, C. (2019) QuatNet: Quaternion-Based Head Pose Estimation with Multiregression Loss. *IEEE Transactions on Multimedia*, **21**, 1035-1046. https://doi.org/10.1109/tmm.2018.2866770

[30] Prokudin, S., Gehler, P. and Nowozin, S. (2018) Deep Directional Statistics: Pose Estimation with Uncertainty Quantification. *Computer Vision-ECCV* 2018, Munich, 8-14 September 2018, 542-559. https://doi.org/10.1007/978-3-030-01240-3_33

[31] Lee, T. (2018) Bayesian Attitude Estimation with the Matrix Fisher Distribution on SO(3). *IEEE Transactions on Automatic Control*, **63**, 3377-3392. https://doi.org/10.1109/tac.2018.2797162

[32] Mohlin, D., Sullivan, J. and Bianchi, G. (2020) Probabilistic Orientation Estimation with Matrix Fisher Distributions. *Advances in Neural Information Processing Systems*, **33**, 4884-4893.

[33] Wang, W. and Lee, T. (2020) Matrix Fisher-Gaussian Distribution on $SO(3) \times R^n$ for Attitude Estimation with a Gyro Bias. 2020 *American Control Conference* (*ACC*), Denver, 1-3 July 2020, 4429-4434. https://doi.org/10.23919/acc45564.2020.9147703

[34] Lewis, J.P., Cordner, M. and Fong, N. (2023) Pose Space Deformation: A Unified Approach to Shape Interpolation and Skeleton-Driven Deformation. In: Whitton, Ed., M.C., *Seminal Graphics Papers*: *Pushing the Boundaries*, *Volume* 2. Association for Computing Machinery, New York, 811-818. https://doi.org/10.1145/3596711.3596796

[35] Zheng, M., Yang, H., Huang, D. and Chen, L. (2022) ImFace: A Nonlinear 3D Morphable Face Model with Implicit Neural Representations. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 20311-20320. https://doi.org/10.1109/cvpr52688.2022.01970